

Prácticas de Análisis Estadístico de Datos Geoquímicos con el Programa R

/ DAVID BENAVENTE (1,2,*), MARINA LEAL PALAZÓN (2), ÁLVARO MARTÍNEZ PERAL (2), MARÍA JOSÉ NUEDA (2,3)

(1) Dpto. Ciencias de la Tierra y del Medio Ambiente. Universidad de Alicante. Campus San Vicente del Raspeig s/n, 03690. San Vicente del Raspeig (Alicante)

(2) Facultad de Ciencias, Universidad de Alicante.

INTRODUCCIÓN

Son muchas las materias de carácter geológico como geoquímica, prospección geoquímica, petrología, recursos minerales, mineralogía ambiental, etc., donde el análisis estadístico de datos geoquímicos es una herramienta básica y necesaria. La toma de datos en el campo o en el laboratorio, el análisis, representación e interpretación de datos que proceden de un laboratorio analítico son tareas comunes en la toma de decisiones o explicación de procesos geoquímicos y petrogenéticos, detección de anomalías geoquímicas, procedencia de aguas, etc. En dichos estudios se suelen realizar análisis químicos de suelos, sedimentos y/o aguas, que incluyen más de una docena de elementos (variables) y varios puntos de muestreo (casos).

Para obtener la máxima información posible de la enorme cantidad de datos geoquímicos generados se requiere de la utilización de las diferentes técnicas estadísticas uni-, bi- y multivariante. Sin embargo, los contenidos de estadística en el grado de Geología, en particular, y los grados de Ciencias, en general, (i) se han reducido a los últimos temas de asignaturas básicas de matemáticas, impartidos con poca profundidad, y (ii) se limitan al análisis descriptivo univariante y bivariante. Este hecho refuerza la necesidad de desarrollar un material específico que cubra las necesidades formativas de las asignaturas donde se trabaja con datos geoquímicos. El objeto de estas prácticas, por lo tanto, no es formar a estadísticos, sino usuarios de estadística aplicada a la geoquímica.

En estas prácticas sólo se va a trabajar con los métodos de análisis estadísticos más habituales en geoquímica. En general, éstos se basan en el estudio descriptivo e inferencial de una o varias

variables. No obstante, dependiendo del objeto del estudio, estos métodos se suelen complementar con la metodología de la Regresión para el estudio de una variable en función de otras, el Análisis de Componentes Principales para resumir información y Análisis de Conglomerados para agrupar muestras a partir de la información disponible. Además, también es habitual utilizar el método de ANOVA (Análisis de la Varianza) para comparación de medias entre distintas poblaciones o bien el Análisis Discriminante para identificar nuevas muestras con alguna clasificación predeterminada.

Se ha elegido el programa R [<http://cran.r-project.org/>] para las prácticas por distintas razones: su robustez de cálculo, la posibilidad de manipular datos y funciones y generar gráficos y por ser de libre acceso, lo que fomenta el conocimiento científico abierto. La comunidad de R es muy activa y está constituida por investigadores de contrastado prestigio, que es la que desarrolla, implementa y comparte los paquetes que se utilizan en R. En los últimos años se ha extendido a otros campos como la geoestadística, lo que confiere a R una herramienta muy potente para trabajar con datos geoquímicos georeferenciados.

Los objetivos que se plantean en el presente trabajo son:

- (1) Proporcionar una metodología estructurada de las técnicas estadísticas básicas (qué análisis estadísticos es preciso aplicar a los datos geoquímicos y en qué orden).
- (2) Entender los conceptos básicos de los principales análisis estadísticos (en qué consiste cada análisis estadístico, por qué se utiliza y qué información se obtiene).
- (3) Introducir al alumno en el entorno de

R (adquirir destrezas en programación informática y fomentar el conocimiento científico abierto).

Este tipo de prácticas han sido desarrolladas para las asignaturas de Geoquímica y Prospección Geoquímica del grado de Geología y de Geofísica Aplicada y Prospección Geoquímica de la licenciatura de Ingeniería Geológica, de la Universidad de Alicante.

Aunque estas prácticas se han centrado en el análisis de datos geoquímicos, el material desarrollado se puede utilizar en cualquier materia tanto docente como investigadora donde se realicen análisis de datos.

MATERIALES Y MÉTODOS

Al alumno se le proporcionan apuntes claros y concisos de cada técnica estadística, scripts (archivos de texto con las instrucciones a ejecutar con R) y funciones de R, así como la colección de problemas geoquímicos que constituyen las prácticas a través del *Campus Virtual* y de la web que se ha desarrollado específicamente para ello [<http://web.ua.es/es/lpa/docencia/analisis-estadistico-de-datos-geoquimicos-con-r/indice.html>].

La colección de problemas o ejercicios que constituyen cada práctica comienza con casos sencillos (ej.: web del Davis, 2002):

<http://www.kgs.ku.edu/Mathgeo/Books/Stat/>], incrementando el grado de dificultad y, en algunos ejercicios, combinando la utilización de otros análisis estadísticos estudiados en otras prácticas. La última práctica estará constituida por uno o varios estudios integrales, en los que se utilicen la mayoría de los test estadísticos aprendidos en un mismo ejercicio, siempre enmarcado dentro un estudio

palabras clave: didáctica, multivariante, software libre

key words: didacticism, multivariate, free software

geoquímico (ej.: Aiuppa *et al.*, 2003). Es muy importante trabajar con estudios geoquímicos reales y completos para que el alumno se sienta más identificado con la problemática que plantea el caso de estudio en particular, entienda el uso de cada técnica estadística y la información que le proporciona dicho análisis para entender un proceso geoquímico complejo.

El desarrollo de las prácticas no necesita del conocimiento de programación en *R* y se basa en la ejecución de un script para cada análisis estadístico que previamente se proporciona. Se necesita estar familiarizado con el entorno de *R* y los elementos que lo componen. Se aconseja entender cada parte de los scripts para comprender los pasos que se realizan durante los análisis estadísticos. Esto permitirá modificar cualquier otro script de forma fácil e intuitiva y motivar al alumnado para profundizar en *R*.

Cada script incluye las diferentes instrucciones que se irán ejecutando de forma directa o bien llamando a programas o funciones. En el script también se explica qué realiza cada una de estas instrucciones. Las explicaciones están contenidas en las líneas que comienzan con el símbolo #. Para que el script sea claro y sencillo de trabajar, se ha preferido dejar los comandos esenciales dentro de cada script y crear programas con las funciones o subrutinas de mayor complejidad.

Los scripts de estas prácticas se han desarrollado para que la matriz de datos geoquímicos a la que se van a aplicar los análisis estadísticos se lea desde el archivo *datos.txt*. Se ha elegido esta metodología para facilitar el trabajo al alumno, minimizar los posibles errores a la hora de nombrar o cargar los archivos de datos y evitar incompatibilidades entre diferentes programas. Un archivo de texto plano *txt* lo lee cualquier programa, como *R* o una hoja de cálculo (*EXCEL* o *CALC del OpenOffice*). Para la creación de bases de datos grandes, se propone trabajar con el archivo *datos.txt* desde la hoja de cálculo. Las hojas de cálculo son más versátiles que *R* a la hora de introducir los datos y, normalmente, el alumno sabe utilizarlas. Algunos análisis también podrían ser realizados con la hoja de cálculo en paralelo con *R*, como la estadística descriptiva o las regresiones

simples lineales y no lineales, debido a que el alumno está más familiarizado y puede asimilar mejor el formato de los resultados que aparecen en la consola de *R*.

Los criterios de evaluación se centran principalmente en:

- (1) Trabajo diario en el aula de informática.
- (2) Memoria de prácticas que incluya varios ejercicios (2 ó 3) donde en cada uno de ellos se utilicen todos los análisis estadísticos (ej.: Aiuppa *et al.*, 2003).
- (3) Un pequeño examen práctico en el aula de informática con el objeto de evaluar el manejo de *R* y el grado de entendimiento de cada técnica estadística, es decir, se evalúa la capacidad del alumno para realizar la memoria de prácticas de forma individual. El examen práctico incluye varios métodos estadísticos estudiados, pudiendo el alumno consultar todo el material de las prácticas. Su duración es relativamente corta, aproximadamente media hora, para premiar a aquellos alumnos que más han trabajado y asimilado las prácticas.

CONSIDERACIONES DIDÁCTICAS FINALES.

Los resultados obtenidos en la implantación de este método han sido, según nuestro criterio, muy satisfactorios desde el punto de vista metodológico (enlazar los contenidos teóricos con la asignatura y con prácticas reales), formativo (se han alcanzado los objetivos del aprendizaje establecidos en la memoria del Grado de Geología) y calificativo.

Este tipo de prácticas ha tenido una buena aceptación por parte de los estudiantes. Se valora su utilidad al poner en práctica los conocimientos adquiridos así como su aplicabilidad en otras asignaturas y en salidas profesionales.

AGRADECIMIENTOS

Este trabajo está enmarcado dentro de la Red de Investigación en Docencia Universitaria (2899) y el Grupo de Innovación Tecnológico-Educativa (GITE-11013-UA: Petrología y Geoquímica) de la Universidad de Alicante.

REFERENCIAS

- Aiuppa, A., Bellomo, S., Brusca, L., D'Alessandro, W., Federico, C. (2003): *Natural and anthropogenic factors affecting groundwater quality of an active volcano (Mt. Etna, Italy)*. *Applied Geochemistry*, **18**, 863–882.
- Davis, J.C. (2002): *Statistics and data analysis in geology*. John Wiley & Sons. 638 p.